Comparative and Functional Analysis of Sortase-Dependent Proteins in the Predicted Secretome of *Lactobacillus salivarius* UCC118†

Jan-Peter van Pijkeren,[1,2] Carlos Canchaya,[2] Kieran A. Ryan,[1,2] Yin Li,[1,2] Marcus J. Claesson,[1,2]
Barbara Sheil,[2] Lothar Steidler,[2] Liam O'Mahony,[2] Gerald F. Fitzgerald,[1,2]
Douwe van Sinderen,[1,2] and Paul W. O'Toole[1,2]*

*Department of Microbiology, University College Cork, Cork, Ireland,[1] and Alimentary Pharmabiotic Centre,
University College Cork, Cork, Ireland[2]*

Surface proteins are important factors in the interaction of probiotic and pathogenic bacteria with their environment or host. We performed a comparative bioinformatic analysis of four publicly available *Lactobacillus* genomes and the genome of *Lactobacillus salivarius* subsp. *salivarius* strain UCC118 to identify secreted proteins and those linked to the cell wall. Proteins were identified which were predicted to be anchored by WXL-binding domains, N- or C-terminal anchors, GW repeats, lipoprotein anchors, or LysM-binding domains. We identified 10 sortase-dependent surface proteins in *L. salivarius* UCC118, including three which are homologous to mucus-binding proteins (LSL_0152, LSL_0311, and LSL_1335), a collagen-binding protein homologue (LSL_2020b), two hypothetical proteins (LSL_1838 and LSL_1902b), an enterococcal surface protein homologue (LSL_1085), a salivary agglutinin-binding homologue (LSL_1832b), an epithelial binding protein homologue (LSL_1319), and a proteinase homologue (LSL_1774b). However, two of the genes are gene fragments and four are pseudogenes, suggesting a lack of selection for their function. Two of the 10 genes were not transcribed in vitro, and 1 gene showed a 10-fold increase in transcript level in stationary phase compared to logarithmic phase. The sortase gene was deleted, and three genes encoding sortase-dependent proteins were disrupted. The sortase mutant and one sortase-dependent protein (mucus-binding homologue) mutant showed a significant reduction in adherence to human epithelial cell lines. The genome-wide investigation of surface proteins can thus help our understanding of their roles in host interaction.

Probiotics have been defined as "live microbial food supplements which beneficially affect[s] the host animal by improving its intestinal microbial balance" (24). Many probiotics are members of the genus *Lactobacillus* and are natural inhabitants of the gastrointestinal tract (1). *Lactobacillus salivarius* subsp. *salivarius* strain UCC118 was isolated from the ileal-cecal region of an adult and first described by Dunne et al. (19). It has a spectrum of probiotic features, such as resistance to acid and bile (20), production of a broad-spectrum bacteriocin (23), attenuation of induced arthritis in an interleukin 10 mouse knockout model (52), alleviation of symptoms associated with mild-to-moderate Crohn's disease (39), and adherence to the intestinal mucosa (18).

Interaction of the bacterium with the mucosal surface of the intestine is partly modulated by surface proteins. One subgroup of surface proteins is the sortase-dependent proteins (SDPs). These proteins have a C-terminal motif, LPXTG, which was first described for gram-positive cocci by Fischetti et al. (21). The motif is recognized by the enzyme sortase (SrtA) (59), which cleaves between the threonine and glycine residues and then covalently links the threonine to the amino group of the pentaglycine cell wall cross bridge of the bacterium (41, 42, 59). Apart from SrtA, two other sortase enzymes have been

characterized, namely, SrtB, recognizing an NPQTN sorting motif (40), and SrtC, recognizing a QVPTGV sorting motif (4). However, the majority of sortase-dependent proteins have no assigned function (8). With regard to lactobacilli, only three reports have been published thus far that describe the functional characterization of proteins belonging to this family (12, 47, 49). Buck and colleagues recently used a genome-based approach to identify three proteins that contribute significantly to adhesion of *L. acidophilus* to a human colonic cell line (12).

The genome of *Lactobacillus salivarius* subsp. *salivarius* UCC118 was recently sequenced and annotated in our laboratory (14). The 2.13-Mb genome consists of a 1.83-Mb chromosome, a 242-kb megaplasmid (pMP118), and two smaller plasmids, pSF118-20 and pSF118-44. Other than being subjected to primary annotation, surface proteins were not analyzed. For the present study, we searched the genome of *L. salivarius* strain UCC118 for the presence of sortase gene homologs and genes encoding sortase-dependent proteins. Moreover, we applied a genome-wide survey of cell wall-anchored proteins in the publicly available *Lactobacillus* genomes and compared these data to those for *L. salivarius* strain UCC118. Transcriptional analysis and functional characterization of targeted gene knockout mutants were employed to examine the role in adhesion of the sortase protein and sortase-dependent proteins. The in vitro adhesion data for epithelial cells demonstrate the role of sortase-dependent proteins in epithelial cell adhesion by *L. salivarius* UCC118 and reveal a significant contribution by the LspA protein to this process.

* Corresponding author. Mailing address: Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland. Phone: 353 21 490 3997. Fax: 353 21 490 3101. E-mail: pwotoole@ucc.ie.

TABLE 1. Bacterial strains and plasmids used in this study

Bacterial strain or plasmid	Relevant properties[a]	Reference or source
L. salivarius strains		
UCC118	Ileocecal isolate from a human adult	19
UCC118Δ*srtA*	UCC118 with deletion of the sortase gene (LSL_1606)	This work
UCC118/pORI19::*lspA*	UCC118 integrant LSL_0311 (*lspA*)::pLS002	This work
UCC118/pORI19::*lspB*	UCC118 integrant LSL_1085 (*lspB*)::pLS003	This work
UCC118/pORI19::*lspD*	UCC118 integrant LSL_1838 (*lspD*)::pLS004	This work
UCC118/pORI19::*lacZ*	UCC118 integrant LSL_0376 (*lacZ*)::pLS005	This work
L. lactis strains		
MG1363	Plasmid-free derivative of *L. lactis* subsp. *cremoris* NCDO712	26
LL108	*L. lactis* strain with *repA* gene integrated in chromosome	35
Plasmids		
pORI19	Em[r] Ori[+] RepA[−] *lacZ′* derivative of pORI28	34
pLS001	pORI19 containing flanks of *srtA*	This work
pLS002	pORI19 containing a 974-bp internal gene fragment of *lspA*	This work
pLS003	pORI19 containing a 999-bp internal gene fragment of *lspB*	This work
pLS004	pORI19 containing a 404-bp internal gene fragment of *lspD*	This work
pLS005	pORI19 containing a 1,002-bp internal gene fragment of *lacZ*	This work
pVE6007	Cm[r] Ts derivative of pWV01	37

[a] Em[r], erythromycin resistant; Ori[+], replication origin; RepA[−], lacking replication protein A; Cm[r], chloramphenicol resistant; Ts, temperature sensitive.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The strains and plasmids used in this study are listed in Table 1. *L. salivarius* subsp. *salivarius* strain UCC118 was cultured at 37°C under microaerophilic conditions (5% CO_2) in de Man-Rogosa-Sharpe (MRS) medium (16) (Oxoid Ltd., Basingstoke, Hampshire, United Kingdom). *Lactococcus lactis* MG1363 and *Lactococcus lactis* LL108 were used in this study as plasmid hosts and were cultured at 30°C in M17 broth (58) (Oxoid Ltd., Basingstoke, Hampshire, United Kingdom) containing 0.5% glucose. When necessary, erythromycin (Em) or chloramphenicol (Cm) was supplemented to a final concentration of 5 μg/ml for *L. salivarius* UCC118 and lactococcal strains.

DNA manipulations. The primers used in this study were purchased from MWG Biotech (Ebersberg, Germany) and are listed in Table S1 in the supplemental material. For cloning purposes, *Pwo* polymerase (Roche, Mannheim, Germany) was used for PCR amplification, while for screening purposes, *Taq* polymerase (Bioline, London, United Kingdom) was used. Restriction endonucleases, T4 DNA ligase, and PCR purification kits were purchased from Roche and used according to the manufacturer's recommendations. Ligation products were precipitated using Pellet Paint (Novagen, United Kingdom) prior to transformation. *L. lactis* LL108 was used as a host for pORI19 constructs. Plasmid DNA was isolated from *L. lactis* by using an *Escherichia coli* plasmid purification kit (QIAGEN, Crawley, United Kingdom) adapted for use with lactococci by the incorporation of 20 mg/ml lysozyme (Sigma, St. Louis, MO). Genomic DNA of *L. salivarius* UCC118 was isolated as previously described (23).

L. salivarius UCC118 was transformed by using a modified procedure based on that of Serror et al. (51). Briefly, MRS containing 1.9% (wt/vol) glycine was inoculated (1% [vol/vol]) from a freshly grown overnight culture. Cells were harvested at an optical density at 600 nm between 0.3 and 0.6 by centrifugation at 4°C (20 min at 3,000 × *g*). The pellet was washed twice with buffer (0.5 M sucrose, 7 mM potassium phosphate [pH 7.4], 1 mM $MgCl_2$) and resuspended in 1/100 the culture volume in the same buffer. Glycerol was added at 1/5 the culture volume and mixed, and 50-μl aliquots of cells were transformed with 1 to 5 μl of DNA (1 μg), using the following parameters: voltage, 1.5 kV; resistance, 400 Ω; and capacitance, 25 μF. Upon transformation, 1 ml recovery buffer (MRS containing 20 mM $MgCl_2$ and 2 mM $CaCl_2$) was added, and cells were incubated for 3 hours at 37°C (5% CO_2). Bacteria were plated on selective MRS-agar plates.

Construction of an isogenic sortase mutant. Genomic DNA of *L. salivarius* UCC118 was used as a template for PCR amplification of the 5′- and 3′-end-flanking regions of the sortase gene (LSL_1606), using primer pairs JP144-JP145 and JP146-JP147. The amplicons were joined by splicing by overlap extension (SOE)-PCR using the primer pair JP144-JP147. The resultant 1.6-kb amplicon was digested using BamHI and EcoRI and cloned into pORI19 digested with the same enzymes. The integrity of the obtained transformants was verified by PCR, using primers ORI47 (located on pORI19) and JP144. The resultant plasmid was

named pLS001. Plasmid integrants in *L. salivarius* UCC118 were constructed as described previously (34), with minor modifications. Briefly, *L. salivarius* UCC118 containing pVE6007 was transformed with pLS001 and cultured for 24 h at 37°C (5% CO_2) with Em selection. Subsequently, cells were passaged for 50 generations at 42°C, with selection for pLS001 only, thus allowing integration into the chromosome upon loss of pVE6007. Colonies were screened for sensitivity to Cm in 96-well plates. Genomic DNAs were prepared from Em[r] Cm[s] cultures, and upstream and/or downstream integration was confirmed by PCR, using primer pairs JP166-ORI47 and JP167-ORI48B, respectively. Plasmid integrants upstream, downstream, and where single crossovers occurred both up- and downstream were selected and cultured at 37°C (5% CO_2) without antibiotic selection for at least 50 generations. Ninety-six colonies were randomly selected and screened for an Em[s] phenotype. From 18 Em[s] cultures, genomic DNA was prepared. The occurrence of a double-crossover event was confirmed for one Em[s] culture by PCR amplification using the primer pair JP166-JP167, which flanks the sortase gene.

Disruption of SDP genes and *lacZ* by plasmid integration. The primer pairs JP082-JP083, JP090-JP091, JP190-JP191, and JP076-JP081 were used for PCR amplification of internal gene fragments of LSL_0311 (*lspA*), LSL_1085 (*lspB*), LSL_1838 (*lspD*), and LSL_0376 (*lacZ*), using genomic DNA of *L. salivarius* UCC118 as the template. PCR amplicons of the internal gene fragments of *lspA*, *lspB*, *lspD*, and *lacZ* were digested with EcoRI-HindIII, HindIII-EcoRI, BamHI-EcoRI, and BamHI-EcoRI, respectively, and cloned into pORI19, which had been treated with the same respective restriction endonucleases. These recombinant plasmids were designated pLS002, pLS003, pLS004, and pLS005, respectively. Single-crossover plasmid integrants were obtained as described above and confirmed by PCR, using primer pairs JP070-ORI47, JP065-ORI48B, JP164-ORI48B, and JP092-ORI48B for UCC118/pORI19::*lspA*, UCC118/pORI19::*lspB*, UCC118/pORI19::*lspD*, and UCC118/pORI19::*lacZ*, respectively.

Southern hybridization. Southern hybridization was performed using an ECL hybridization and detection kit (Amersham Biosciences, United Kingdom). Probes were identical to the PCR amplicons of the internal gene fragments described above. The double-crossover deletion was confirmed by using a probe which was complementary to the upstream sequence of the sortase gene. The probe was generated using the primer pair JP144-JP149.

Adhesion assay by viable count method. The colonic cell line C2/bbe1, a differentiated subclone of Caco-2 cells, and the adenocarcinogenic cell line HT29 were used to assess the adhesion abilities of the constructed mutants. C2/bbe1 cells were maintained in Dulbecco's modified Eagle's medium (Sigma, St. Louis, MO) supplemented with 10% (vol/vol) heat-inactivated (10 min at 70°C) bovine serum, nonessential amino acids, and 10 μg/ml human transferrin. HT29 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% (vol/vol) heat-inactivated (10 min at 70°C) bovine serum. Adhesion assays were performed essentially as described previously (28). Briefly, six-well plates were seeded with 1×10^6 cells/well, and upon confluence, C2/bbe1 cells were

maintained for 21 days or HT29 cells were maintained for 12 to 15 days to allow the cells to completely differentiate. Prior to the assay, monolayers were washed twice with phosphate-buffered saline (PBS). Bacterial overnight cultures were washed once with PBS, adjusted to an optical density at 600 nm of 1.0, and diluted 10-fold in PBS to reach ~1×10^8 CFU/ml, as determined by plate counts on MRS-agar. One milliliter of the bacterial suspension was added to the washed monolayers (bacterial cell/epithelial cell ratio of ~50:1) and incubated for 30 min at 37°C (5% CO_2). Monolayers were washed five times with PBS to remove unbound bacteria. Adherent cells were removed by scraping, serially diluted in PBS, and plated on MRS-agar plates. Adhesion was expressed relative to that of the *L. salivarius* UCC118 wild-type strain. Adhesion assays were performed in duplicate in three independent experiments.

Adhesion assay by semiquantitative real-time PCR. Twenty-four-well plates were seeded with 2×10^5 cells/well and maintained until the cells were fully differentiated, as described above. Monolayers and bacterial suspensions were prepared as described above, and 250 μl of bacterial suspension was added per well. Following incubation, washed monolayers were scraped off, resuspended in 250 μl PBS, and combined with 500 μl sterile distilled H_2O and 1 g 0.1-mm zirconia-silica beads (Biospec Products, Bartlesville, OK). Total lysis was achieved by bead beating for 1 min at maximum speed in a mini-bead beater (Stratech, United Kingdom). Cellular debris was pelleted by centrifugation for 2 min at 13,000 × *g*. Twenty microliters of supernatant was combined with 980 μl sterile distilled H_2O to dilute PCR inhibitors, and 1 μl was used as template DNA for real-time PCR. Real-time PCR was performed using the ABI7000 system (Applied Biosystems, Foster City, CA). PCR master mix was purchased from Biogene (Kimbolton, United Kingdom). A chromosomally located pseudogene (LSL_1319) was chosen as a target for amplification using the primer pair CC01-CC02 (see Table S1 in the supplemental material). A target was chosen from the chromosome instead of any of the resident plasmids to avoid complications that might occur from variations in plasmid copy number. Quantification of adherent bacteria was done by a standard curve method. Adhesion was performed in triplicate in three independent experiments.

Gene expression analysis. To examine the expression of the genes encoding seven identified sortase-dependent proteins in stationary-phase cells compared to that in logarithmic-phase cells, total RNA from cells in both growth phases was isolated using an RNA-easy kit (Ambion, Cambridgeshire, United Kingdom). The Improm-II reverse transcriptase enzyme (Promega, Madison, WI) was used to prepare cDNA according to the manufacturer's recommendations. Primers for real-time PCR were designed using the web-based tool Primer3 (50). For the three pseudogenes, primers were designed to target regions upstream and downstream of the internal stop codon (see Table S1 in the supplemental material). Real-time PCR was performed as described above. Gene expression levels were expressed relative to that of the 16S rRNA gene, as previously described (46). Gene expression analysis of both growth phases was investigated in three independent experiments.

Sequence analysis. The *L. salivarius* genome sequence (accession no. CP000233 [chromosome] and CP000234 [pMP118 megaplasmid]) has recently been determined (14). The genome sequences of *L. plantarum* WCFS1 (AL935263) (31), *L. johnsonii* NCC 533 (AE017198) (48), *L. acidophilus* NCFM (CP000033) (2), and *L. sakei* 23K (CR936503) (13) were also analyzed. Sequence analysis was performed with BLAST (3). Signal peptide prediction and cleavage site prediction were performed with SignalP3.0 (6). Transmembrane helices were predicted using the TMHMM server (32). The presence of LysM domains, peptidoglycan-binding domains, and choline-binding domains was determined by screening against the Pfam database (5), and results were filtered using an E value cutoff of $<1 \times 10^{-5}$. Lipoprotein predictions were performed as previously described (57). Sequences were searched for a WXL-binding domain by using the search string [LI]TW[TS]L, and the results were screened manually to determine the location of the motif within the sequence. Protein or DNA repeats were identified by using the programs Dotter (53) and RADAR (27). Sequences containing repetitive regions were screened manually for the presence of GW residues. Sortase substrates were identified by manual screening and a hidden Markov model (8).

Statistical analysis. Student's *t* test was employed to investigate statistical differences. Samples with *P* values of <0.05 were considered statistically different.

RESULTS

Identification of cell wall-anchored proteins. Using the annotated genome sequence of *L. salivarius* UCC118, a bioinfor-

matic approach was employed to identify secreted proteins, including those predicted to be cell wall anchored. The results were compared to the data from parallel analyses of the genomes of *L. plantarum* WCFS1, *L. acidophilus* NCFM, *L. johnsonii* NCC 533, and *L. sakei* 23K (Table 2).

L. salivarius UCC118 possesses the second-largest genome (2.13 Mb) of the fully sequenced lactobacilli and is the only sequenced *Lactobacillus* strain harboring a megaplasmid (14). Using SignalP3.0, we identified 119 proteins predicted to be secreted, the majority (108) of which are encoded by the chromosome. Eight are encoded by the megaplasmid (pMP118), two are encoded by the 44-kb plasmid pSF118-44, and one is encoded by the 20-kb plasmid pSF118-20. Deduced products of an additional five pseudogenes were predicted to be secreted, with two encoded by the chromosome and three encoded by pMP118. Of the 119 proteins, 44 were predicted to be cleaved by signal peptidases I and 3 were predicted to be cleaved by signal peptidase II, and thus the majority of secreted proteins will remain associated with the cell membrane (Table 2). For other *Lactobacillus* genomes analyzed, the distributions of identity levels for existing database entries were similar, with a preponderance of proteins with values centrally distributed around identities of 30 to 60% (Table 2). The exception was *L. johnsonii*, for which a larger proportion of secreted and anchored proteins (see below) displayed significant identity to database entries.

We used a combination of manual inspection of proteins predicted to be secreted and the hidden Markov model of Boekhorst et al. (8) to identify sortase substrates. The hidden Markov model was also used to search the genomes of *L. sakei* and *L. acidophilus*, whereas sortase substrates for *L. plantarum* and *L. johnsonii* have been described previously (8). We thus identified 10 proteins containing sortase substrates in *L. salivarius* (Table 3), one of which is encoded by the previously characterized 44-kb plasmid pSF118-44 (22). Four of these proteins are encoded by pMP118, and five are encoded by chromosomal genes. Two SDPs are theoretical products of gene fragments, and four theoretical proteins were derived from pseudogenes caused by interruption with an internal stop codon or a frameshift.

One of the pseudogenes, designated LSL_0152, encodes a protein which shares 30% identity with the mucus-binding protein Mub of *L. reuteri* (49). LSL_0152 is interrupted by a stop codon (TGA) at nucleotide position 499. Another pseudogene, LSL_1319, shows 21% identity to the R28 protein of *Streptococcus pyogenes* (54), which is involved in binding to epithelial cells. The DNA sequence of LSL_1319 is interrupted by a stop codon (TAA) at nucleotide position 667. The third pseudogene, LSL_2020b, is located on pSF118-44. The encoded protein shares 25% identity with the collagen adhesin of *Staphylococcus aureus* (44). LSL_2020b is interrupted after 1,938 base pairs by the stop codon TAA. For these pseudogenes, there is no evidence of a second ribosome binding site with a start codon, which could lead to translation of the distal fragment of the gene. Whereas the other pseudogenes are disrupted by a stop codon, we identified a pseudogene (LSL_1774b) with a frameshift in its sequence which introduced a stop codon. The product of LSL_1774b is homologous (>32% identity) to a 1,480-amino-acid proteinase (PrtR) of the human isolate *L. rhamnosus* BGT10 (43). Apart from the pseudogenes, two

TABLE 2. Genome-wide survey of cell wall-anchored proteins in *L. salivarius* UCC118 and comparison with available *Lactobacillus* genomes

Protein feature	*L. salivarius* UCC118				*L. plantarum* WCFS1[a]				*L. johnsonii* NCC 533[a]				*L. acidophilus* NCFM				*L. sakei* 23K			
	No. of proteins[b]	No. of proteins with BLAST-NR % identity cutoff of:			No. of proteins[b]	No. of proteins with BLAST-NR % identity cutoff of:			No. of proteins[b]	No. of proteins with BLAST-NR % identity cutoff of:			No. of proteins[b]	No. of proteins with BLAST-NR % identity cutoff of:			No. of proteins[b]	No. of proteins with BLAST-NR % identity cutoff of:		
		>60	30–60	<30		>60	30–60	<30		>60	30–60	<30		>60	30–60	<30		>60	30–60	<30
Export features																				
Signal sequence[c]	124	27	70	27	217	30	151	36	128	93	27	8	173	60	78	35	144	35	78	31
SPase I cleavage[d]	44	6	27	11	86	12	51	23	31	21	9	1	56	10	28	18	50	10	25	15
Cell-wall-anchored proteins																				
N- or C-terminally anchored proteins	80	18	41	21	131	20	99	12	97	72	18	7	117	53	47	17	94	24	54	16
LPXTG anchors[e]	10	0	5	5	27	0	15	12	16	3	9	4	12	1	6	5	4	0	2	2
Lipoprotein anchors[f]	3	1	2	0	3	0	3	0	1	1	0	0	5	3	0	2	2	0	2	0
Choline binding domain[g]	0				0				0				0				0			
Peptidoglycan binding domain[g]	0				1	0	1	0	1	0	1	0	0				0			
GW repeats[h]	4	1	2	1	11	3	5	3	1	1	0	0	3	0	3	0	3	1	1	1
LysM domain[g]	9	0	6	3	11	6	5	0	1	0	1	0	1	0	1	0	4	1	2	1
WXL domain[i]	1	0	1	0	6	0	3	3	0				0				9	0	7	2

[a] Proteins with LPXTG anchors were previously identified by Boekhorst et al. (8).
[b] Values tabulated include pseudogenes.
[c] Signal sequence prediction was done with the hidden Markov model in SignalP3.0, with *P* values of >0.95 as the cutoff.
[d] Cleavage site prediction was done with the neural network model in SignalP3.0, with C_{max} values of >0.52 and Y_{max} values of >0.32 as cutoffs.
[e] LPXTG anchors were predicted by the hidden Markov model for sortase substrates (8).
[f] Lipoprotein prediction was done as described by Sutcliffe and Harrington (57).
[g] From the Pfam database, with a cutoff E value of ≤10^{-5}.
[h] Manual screening for the presence of GW residues in repetitive regions was performed.
[i] Prediction based on the presence of the [LI]TW[TS]L motif in the C-terminal sequence.

TABLE 3. Putative sortase-dependent proteins of *L. salivarius* UCC118 with relevant properties

Gene identifier	Annotation	Cleavage motif	Gene location	Size (aa)[a]	Predicted protein size (kDa)[b]	Top BLAST hit (accession no.), organism[c]	E value
LSL_0152	*mbp-1*	LPQTG	Chromosome	166, 2,885	NA	Hypothetical protein (NP_964406), *L. johnsonii* NCC 533	2e−91
LSL_0311	*lspA*	LPQTG	Chromosome	1,209	131	Hypothetical surface protein (ZP_00874951), *S. suis* 89/1591	1e−48
LSL_1085	*lspB*	LPQMG	Chromosome	827	87	Enterococcal surface protein (AAQ89938), *E. faecium*	1e−04
LSL_1319	*rlpA*	LPQTG	Chromosome	222, 1,043	NA	Hypothetical protein (NP_965634), *L. johnsonii* NCC 533	3e−77
LSL_1335	*lspC*	LPQTG	Chromosome	785	88	Hypothetical protein (NP_964510), *L. johnsonii* NCC 533	2e−16
LSL_1838	*lspD*	LPQTG	pMP118	493	52	Hypothetical protein (XP_364726), *Magnaporthe grisea* 70-15	5e−13
LSL_1774b	*prtP*	LPQTG	pMP118	842, 685	NA	PrtP precursor (AAV43331), *L. acidophilus* NCFM	0.0
LSL_1832b	*sapA*	LPQMG	pMP118	**775**	85	Streptococcal surface protein A precursor (AAC44101), *S. gordonii*	3e−75
LSL_1902b	Hypothetical	LPQTG	pMP118	**49**	5.3	Hypothetical protein (XP_500168), *Yarrowia lipolytica*	5.5
LSL_2020b	*cna*	LPQTG	pSF118-44	646, 325	NA	Collagen binding precursor (ABA12809), *L. paracasei* subsp. *paracasei*	0.0

[a] For pseudogenes, the numbers of amino acids upstream and downstream of the internal stop codon, respectively, are indicated. Numbers in bold indicate gene fragments.

[b] NA, not applicable.

[c] BLAST hits were generated by comparing the six-frame translation output of the target nucleotide sequence to a protein database, using the BLAST-X algorithm.

gene fragments harboring a sortase recognition sequence were also identified (Table 3). Both gene fragments are located on pMP118. LSL_1832b is a 2.3-kb gene fragment whose derived amino acid sequence harbors an LPQMG sortase recognition motif. The fragment is homologous (>17% identity) to the C-terminal region of a 1,575-amino-acid salivary agglutinin-binding protein of *Streptococcus gordonii* (17). The smallest gene fragment harboring a sortase recognition motif is LSL_1902b (147 base pairs). It has no homology with proteins in the nonredundant BLAST database. Apart from the six interrupted/partial genes containing sortase recognition motifs, we identified four predicted sortase-anchored proteins which are intact, designated *Lactobacillus* surface proteins A, B, C, and D (LspA, LspB, LspC, and LspD, respectively) (Table 3).

LspA (LSL_0311) is a 1,209-amino-acid protein which contains seven repeats of 79 amino acids (R1 to R7) (Fig. 1). R1 and R7 are the least conserved repeats, sharing 73% identity, whereas R2 to R6 are more conserved, sharing >92% identity. Pfam analysis revealed that each of these repeats is similar to mucus-binding domains (PF06458), with E values ranging from 10^{-1} to 10^{-6} but with all scores being above the gathering threshold. BLAST-NR searches did not reveal homology to a functionally characterized protein, since the closest homologue is a hypothetical protein of *Streptococcus suis* (ZP_00874951), as shown in Fig. 1. LspB (LSL_1085) is an 827-amino-acid protein (Fig. 1) containing an LPQMG cleavage motif. Three 13-amino-acid repeats were identified at the C-terminal end of the protein. The repeats are 100% identical, and Pfam analysis revealed no predicted function. The top BLAST hit for LspB is an enterococcal surface protein (Esp) of *Enterococcus faecium* (AAQ89938) which has no assigned function (Fig. 1). LspC (LSL_1335) is 785 amino acids in size and has four repeats of 97 amino acids (Fig. 1). There is over 98% identity among

these repeats, and their sequences are similar to those of mucus-binding domains, as predicted by Pfam analysis, with E values ranging between 10^{-3} and 10^{-4} but with all scores being above the gathering threshold. It is homologous to the 3,269-amino-acid mucus-binding protein (Mub [AAF25576]) previously characterized in *L. reuteri* (49). This protein has two types of repeats. One set of repeats is divergent, with 15 to 85% identity, whereas the second set of repeats is conserved, displaying >91% identity. Both types have been shown to be involved in binding to mucin components (49). The four repeats of LspC show a higher sequence identity to the diverse repeats of Mub (13% identity), whereas there is very low identity (5% identity) to the conserved repeats of Mub. LspD (LSL_1838) is encoded by pMP118 and consists of 493 amino acids (Fig. 1). No repeats were identified, and the top BLAST hit is a hypothetical protein of the fungus *Magnaporthe grisea* (15.4% identity). Similar homology was noted for a sortase-dependent hypothetical protein of *Streptococcus agalactiae* (NP_735436). Since LspD is plasmid encoded, it is noteworthy that there is 15% homology to PrgA, a hypothetical surface exclusion protein of *Enterococcus faecalis* (45). Surface exclusion proteins block the conjugative transfer of plasmids to cells bearing identical or closely related plasmids (15).

The hidden Markov model was also used to search the genomes of *L. sakei* and *L. acidophilus*, but no additional sortase substrates were identified (see Table S5 in the supplemental material).

We identified three lipoprotein sequences in *L. salivarius* UCC118 (see Table S5 in the supplemental material). One of the proteins (LSL_0953) is a hypothetical protein with no clear function; it has homology to a phage-like protein but is not clustered in any of the four phage-associated regions (Sal1 through Sal4) in *L. salivarius* UCC118 (60). The second lipoprotein, LSL_0969, is also a hypothetical protein, whereas
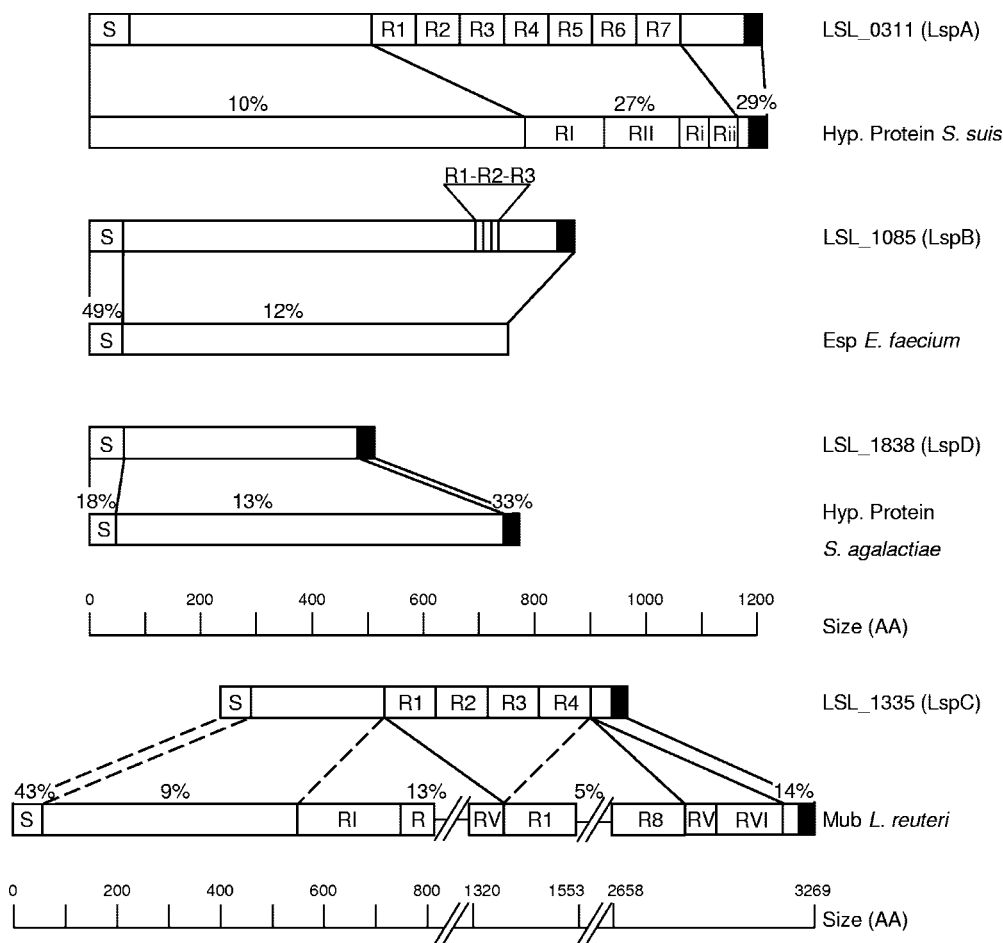
FIG. 1. Diagrammatic representation of the molecular organization of the intact sortase-dependent proteins of *L. salivarius* UCC118 and their homologues. Where possible, proteins were compared in their signal sequences (S), repetitive regions (R), and C-terminal regions (black boxes) containing the transmembrane helix, cell wall anchor, and positively charged tail. A different subset of repeats within a protein sequence is indicated by different numbering (i.e., numerical versus Roman numerals). The comparison of the repetitive region of LspC with the type I repeats of Mub (RI to RVI) is indicated by dashed lines, whereas the comparison to the type II repeats of Mub (R1 to R8) is indicated by solid lines. The sequence identities between the different protein regions are expressed as percentages of identity.

we annotated LSL_1445 as a glutamine-binding protein (see Table S5 in the supplemental material). Twenty-five lipoproteins were previously identified in *L. plantarum* (31), but when we applied a refined search string for the identification of lipoproteins in gram-positive bacteria, as described previously by Sutcliffe and Harrington (57), only three sequences were identified as lipoproteins. Five lipoproteins were identified in *L. acidophilus*, two were identified in *L. sakei*, and *L. johnsonii* appeared to have one lipoprotein (see Table S5 in the supplemental material).

We searched the Pfam database to identify proteins which can be anchored to the cell surface by choline-anchoring domains (25). Results were filtered using an E value cutoff of <1 × 10^{-5}. No sequences in the screened *Lactobacillus* genomes were found to contain a choline-binding domain with an E value below the set cutoff value. Similarly, no proteins in *L. salivarius* UCC118 were found to contain a peptidoglycan-binding domain with a value below the cutoff. Based on the set cutoff, *L. johnsonii* and *L. plantarum* each have a single protein

harboring a peptidoglycan-binding domain (see Table S5 in the supplemental material).

GW repeats have been shown to mediate binding to the cell envelope (38). Four proteins were identified in *L. salivarius* UCC118 that contain repetitive sequences containing GW residues (see Table S5 in the supplemental material). Two proteins with GW repeats are phage related: LSL_0295 is a protein with a hypothetical function and is part of Sal2, an inducible phage in *L. salivarius* UCC118 (60), and LSL_0783 was annotated as a phage terminase and was part of a phage remnant (60). LSL_0982 is predicted to encode a glycosyltransferase in exopolysaccharide gene cluster 1 (14). LSL_1266 was annotated as RNase BN. Comparison with the Cluster of Orthologous Genes database indicated that LSL_1266 is a membrane protein (see Table S5 in the supplemental material). Meanwhile, 11 proteins with GW repeats were identified in *L. plantarum*, 1 was identified in *L. johnsonii*, and 3 each were identified in *L. acidophilus* and *L. sakei* (see Table S5 in the supplemental material).

Proteins can be anchored to the cell envelope by LysM domains, which bind to the peptidoglycan in the bacterial cell wall (55). We identified nine proteins in *L. salivarius* UCC118 with such a domain (see Table S5 in the supplemental material). Two proteins, LSL_0304 and LSL_0805, are phage related and were both annotated as lysozyme. LSL_0304 belongs to Sal2, whereas LSL_0805 belongs to Sal1, a phage remnant (60). Three LysM-type proteins were annotated as hypothetical proteins (LSL_0090, LSL_0901, and LSL_1267), and three proteins were annotated as peptidoglycan binding proteins (LSL_1034, LSL_1036, and LSL_1371). One protein harboring a LysM domain has sequence similarity to a teichoic acid translocation ATP-binding protein (TagH; LSL_0373). The sequence identity with TagH is mainly at the N-terminal end of LSL_0373, whereas the LysM domain is located at the C-terminal end (see Table S5 in the supplemental material). Furthermore, we identified 11 proteins with a LysM domain in *L. plantarum*, 1 each in *L. johnsonii* and *L. acidophilus*, and 4 in *L. sakei* (see Table S5 in the supplemental material).

Kleerebezem and coworkers (31) identified a novel C-terminal WXL domain which they proposed could be a binding domain for the cell envelope, and 19 proteins containing this domain were identified in *L. plantarum*. Chaillou et al. (13) identified 15 proteins containing a YXXT(L/I)TW(T/S)L motif in *L. sakei*. However, when we applied this search motif, no proteins with this motif could be identified in *L. sakei*. The motif was modified to [LI]TW[TS]L, and this search returned nine proteins for *L. sakei*. No C-terminal WXL motif could be identified in the *L. sakei* proteins LSA0611 and LSA1731 (see Table S5 in the supplemental material). An [LI]TW[TS]L motif was identified in one protein in *L. salivarius* (LSL_1295) (see Table S5 in the supplemental material), which is homologous to a neopullulanase, whereas the motif was identified in six proteins of *L. plantarum*. No proteins with this motif were identified in *L. acidophilus* or *L. johnsonii*.

Construction of an isogenic sortase mutant. Previous studies targeting the sortase gene have shown that sortase-dependent proteins play a role in adhesion and virulence in a range of organisms (7, 9, 29, 30, 33). In order to investigate whether a sortase-dependent protein(s) in *L. salivarius* UCC118 is involved in adhesion, we constructed a mutant strain lacking the sortase gene (LSL_1606). The small size of sortase did not allow us to disrupt the gene by plasmid integration, and we therefore opted for a gene deletion, using a double-crossover strategy. Upstream and downstream flanking regions of 772 bp and 818 bp, respectively, were amplified. The upstream flanking amplicon includes the first 13 codons of the sortase gene, whereas the downstream flanking amplicon includes the last 3 codons. Both flanking amplicons were joined by SOE-PCR and cloned into pORI19. The resultant recombinant plasmid, pLS001, was transformed into *L. salivarius* UCC118 harboring pVE6007, and a double-crossover mutant was obtained as described in Materials and Methods. The deletion of the sortase gene in strain UCC118 was verified by Southern hybridization (Fig. 2). A PCR using wild-type genomic DNA with the primer pair JP144-JP149 resulted in a 1.1-kb amplicon which was used as a probe. Genomic DNAs of both the wild-type strain and the sortase mutant were digested with XhoI. The hybridization patterns showed bands of 5.8 kb and 5.1 kb for the wild-type and mutant strains, respectively (Fig. 2). An XbaI-XhoI double
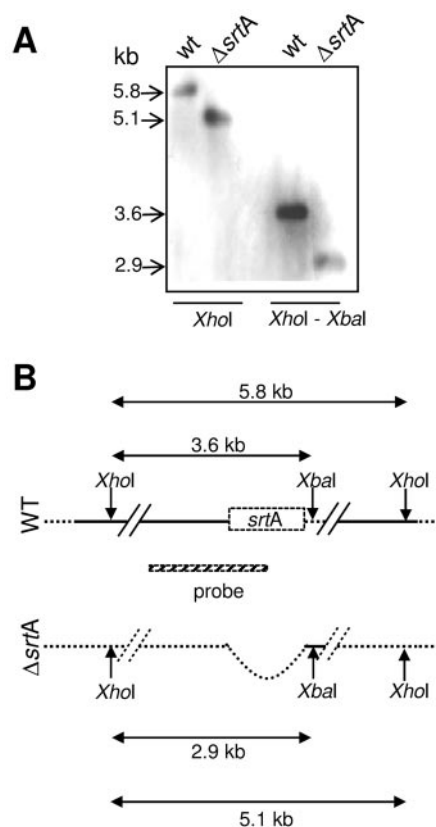

FIG. 2. Verification of the genome structure of a sortase gene deletion mutant. (A) Southern hybridization. The fragments expected following digestion are indicated by arrows, and fragment sizes are indicated in kilobase pairs. (B) Schematic overview. The sortase gene is indicated as a box, whereas the probe is indicated as a hatched box.

digest produced bands of 3.6 kb and 2.9 kb for the wild-type and mutant strains, respectively, confirming the deletion of sortase. The strain lacking the sortase gene was designated UCC118Δ*srtA*.

The sortase mutant has reduced adhesion to epithelial cells. Following the construction of UCC118Δ*srtA*, we tested the strain for adhesion to intestinal epithelial cells. UCC118Δ*srtA* adhered significantly less to HT29 cells ($P = 0.04$) than the wild-type strain did (Fig. 3A). We also employed a semiquantitative real-time PCR method to validate the viable count method (Fig. 3B). The adhesion of the UCC118Δ*srtA* mutant was also significantly reduced ($P = 0.04$) as measured by this method, at 61% of the level of the wild-type strain. The adhesion of UCC118Δ*srtA* to Caco C2 cells was also reduced significantly (68%; $P = 0.007$) compared to that of the wild-type strain, as determined by real-time PCR, but this reduction was less than that observed for the sortase gene mutant grown on HT29 cells. Collectively, these data indicate that one or more sortase-dependent proteins are involved in adhesion to human epithelial cells.

Transcriptional analysis of sortase-dependent proteins. We employed endpoint reverse transcription-PCR (RT-PCR) to test if genes encoding sortase-dependent proteins in *L. salivarius* UCC118 were expressed in vitro when the strain was cultured in MRS broth. RNA was prepared from stationary-phase
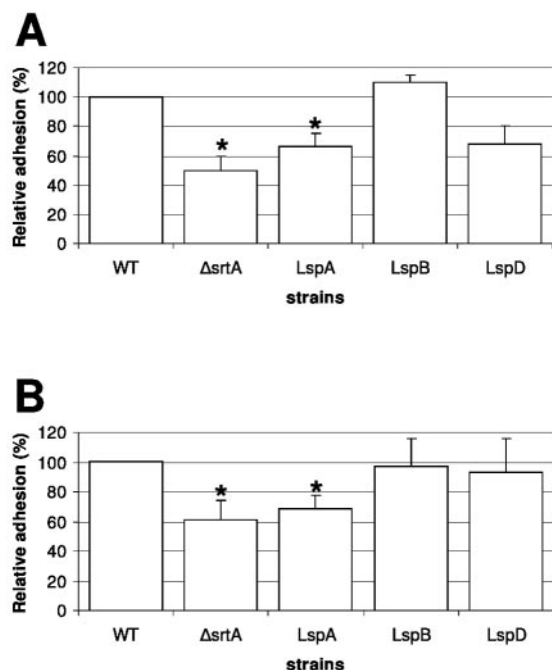
FIG. 3. Adhesion to HT29 cells of the *L. salivarius* UCC118 wild type and mutants lacking the indicated proteins, as determined by the viable count method (A) and semiquantitative real-time PCR (B). The results shown are averages of three independent experiments. Percentages of adhesion are expressed as relative adherence compared to that of the wild-type strain, and the error bars represent standard errors of the means. Statistically significant differences ($P < 0.05$) were determined by Student's *t* test and are indicated with asterisks.
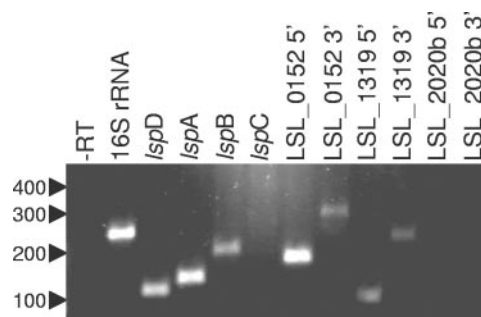

FIG. 4. Expression analysis of genes encoding sortase-dependent proteins. PCR was performed on cDNA prepared from stationary-phase cells grown in MRS broth. Arrows indicate sizes, in base pairs. Genes are indicated above the lanes. Gene labels with 5′ or 3′ suffixes indicate that expression was tested upstream or downstream of the internal stop codon.

cells and reverse transcribed. For each target gene, internal primers were designed, and for the three pseudogenes, primers were designed upstream and downstream of the internal stop codon (see Table S1 in the supplemental material). After 50 cycles of PCR, no gene expression was detected for the chromosomally located gene *lspC* and the pseudogene LSL_2020b, which is carried on the previously described 44-kb plasmid pSF118-44 (22). The remainder of the sortase-dependent proteins were expressed (Fig. 4).

Previously, it was reported that the adhesion of *L. salivarius* UCC118 to HT29 cells is growth phase dependent (18), with a significant increase upon entry into stationary phase. We therefore investigated whether there was differential gene expression of sortase-dependent proteins in the two growth phases by performing real-time PCR, using RNA isolated from cells in the respective growth phases. All of the genes except the pseudogene LSL_1319 were transcribed at higher levels in stationary phase than in logarithmic phase (Table 4). The pseudogene LSL_0152 was transcribed at 2- to 2.5-fold higher levels in stationary phase than in logarithmic phase, while the transcription of *lspB* and *lspD* increased dramatically (Table 4). No gene expression was detected for *lspC* and LSL_2020b in logarithmic-phase cells (data not shown).

Role in adherence of LspA, LspB, and LspD. Transcriptional analysis showed that the expression of *lspC* was not detected after 50 cycles of PCR, and we therefore omitted this gene as a target for disruption. Internal gene fragments of *lspA*, *lspB*,

and *lspD* were cloned into pORI19, and the resulting constructs were designated pLS002, pLS003, and pLS004, respectively. The genes for *lspA*, *lspB*, and *lspD* were thus disrupted by single-crossover plasmid integration, as described in Materials and Methods. Gene disruption was verified by PCR analysis, using genomic DNA as the template, with primers internal to the Em gene of the integrated plasmid and primers located upstream or downstream of the target gene (data not shown). Integrants of the individual target genes were analyzed by Southern hybridization to confirm the genomic arrangement (data not shown). The gene encoding LspD is located on the 242-kb megaplasmid pMP118 in *L. salivarius* UCC118, and it has been reported that this plasmid is present at four or five copies per cell (14). Southern hybridization confirmed the disruption of *lspD* in all copies of pMP118 (data not shown).

L. salivarius UCC118/pORI19::*lspA* adhered significantly less to HT29 cells ($P = 0.001$) than the wild-type strain did (Fig. 3A). A control *lacZ* mutant constructed by pORI19 integration adhered at 97.1% ± 4.4% of the level of the wild-type strain, showing that growth in Em alone was not responsible for lowered adhesion (data not shown). A reduction in adherence was also recorded for *L. salivarius* UCC118/pORI19::*lspD* (Fig. 3), but it was not significant ($P = 0.23$). No significant difference in adhesion was recorded for UCC118/pORI19::*lspB* from that of the wild-type strain ($P =$

TABLE 4. Differential expression of genes encoding sortase-dependent proteins in *Lactobacillus salivarius* UCC118

Gene[a]	Up-regulation (fold [mean ± SEM]) in stationary-phase cells compared to log-phase cells[b]
lspA	1.7 ± 0.2
lspB	5.2 ± 2.5
lspD	10.5 ± 3.0
LSL_0152 (1)	2.0 ± 0.6
LSL_0152 (2)	2.5 ± 0.4
LSL_1319 (1)	1.2 ± 0.1
LSL_1319 (2)	0.8 ± 0.3

[a] Numbers in parentheses following gene names indicate expression upstream of the internal stop codon (1) or downstream of the internal stop codon (2).
[b] Values given were determined relative to the constitutively expressed 16S rRNA gene. Data shown are averages from three independent experiments.

0.56). These data were corroborated by the semiquantitative PCR assay (Fig. 3B), by which statistical significance was detected only for the adhesion reduction of the *lspA* knockout strain. Adhesion to Caco C2 cells was also significantly reduced for strain UCC118/pORI19::*lspA* (77%; $P = 0.009$) but was not significantly reduced for the *lspB* and *lspD* mutants (92% and 94%, respectively, as determined by real-time PCR [data not shown]).

DISCUSSION

The chromosome of *L. salivarius* UCC118 potentially encodes 108 secreted proteins, which comprise 6.2% of the chromosomally located open reading frames (ORFs). This is lower than the proportions for *L. acidophilus*, *L. plantarum*, *L. sakei*, and *L. johnsonii*, all of which devote >7% of their coding capacities to secreted proteins. Interestingly, there are only eight ORFs identified on pMP118 of *L. salivarius* whose products are predicted to be secreted, which is 1.9% of the plasmid-located ORFs. Six of these encode proteins with hypothetical functions, including a protein with an LPQTG sorting motif (*lspD*) and one putative thioredoxin. The two secreted proteins encoded by pMP118 with assigned functions are an oligopeptide binding protein and an amino acid transporter. The plasmids pSF118-20 and pSF118-44 contribute little to the predicted *L. salivarius* UCC118 secretome, encoding one and two secreted proteins, respectively.

Ten proteins were identified in *L. salivarius* UCC118 as sortase substrates by manual screening, and searching the genome with a hidden Markov model (8) did not identify additional potential SDPs. Among these 10 proteins, two sortase substrates were encoded by gene fragments and four were encoded by pseudogenes interrupted by a single stop codon or frameshift. This could be related to genome decay and adaptation of the bacterium to its environment, as previously proposed for *Streptococcus thermophilus* (10). This bacterium has lost most of its ancestral virulence genes, some of which were sortase dependent, and has adapted to a new environment (milk) in which these virulence-related genes are no longer required. One parallel example in *L. salivarius* may be the gene fragment LSL_1832b, which encodes a homolog of a salivary agglutinin-binding protein found in streptococcal bacteria, which plays a role in tooth decay (17). *L. salivarius* may be found in the oral cavity and saliva (1), but strain UCC118 was isolated from the ileal-cecal region (19), and it could be argued that this particular strain has lost a functional salivary agglutinin-binding protein because of a lack of selection. Other examples of nonfunctional adhesins encoded by pseudogenes in *L. salivarius* UCC118 are LSL_2020b (collagen adhesin) and LSL_0152, a large mucus-binding protein that is orthologous to mucus-binding proteins found in many other species of *Lactobacillus* (12, 49).

Interestingly, with the exception of LSL_0152 and LSL_1774b, the remainder of the sortase-dependent proteins have GC contents which are, on average, 5% higher than the average genomic GC content. Two genes in UCC118, *lspA* and *lspB*, display GC contents of 40.0% and 39.5%, respectively, which differ from the average genomic GC content (33%) by more than twice the standard deviation. Thus, it could be argued that these genes were acquired via horizontal gene transfer. In examining the GC contents of the SDPs in the other *Lactobacillus* genomes investigated in this study, we noted that the GC contents were, on average, 3% higher than the average GC contents of the respective genomes, which suggests that many SDPs may have been acquired by horizontal gene transfer. Alternatively, there may be selection for higher GC contents in surface protein-encoding genes.

Transcriptional analysis showed that *lspC* and the pSF118-44-carried pseudogene LSL_2020b were not transcribed in vitro. It remains possible, however, that LspC plays a role in the microbe-host interaction. For example, a subset of *L. plantarum* genes are specifically induced in the murine gastrointestinal tract compared to what occurs under in vitro conditions, including two genes coding for sortase-dependent proteins (11). In *L. salivarius* UCC118, all genes encoding SDPs were upregulated in the stationary growth phase compared to the logarithmic phase, with the exception of the transcription of LSL_1319 downstream of the internal stop codon. The transcription of *lspD* increased 10-fold. Since *lspD* is on a plasmid, it is possible that the abundance of transcript is because of plasmid copy variation in the two growth phases, but this remains to be determined. In *L. plantarum*, an *agr*-like two-component regulatory system was identified which regulated adherence to a glass surface in a growth-phase-dependent manner (56). A homologous system was not annotated in *L. salivarius* UCC118, and the mechanism for growth phase dependent gene regulation is currently unclear.

By individual gene targeting, we showed that UCC118/pORI19::*lspA* had a significant reduction in adherence. There is a predicted terminator sequence located downstream of *lspA*, and therefore the reduced binding to epithelial cells is unlikely to be due to polar effects of the pORI integration on downstream genes. The *lspA* mutant adhered only 15% better than UCC118 lacking sortase, illustrating the importance of LspA in this adhesion model. LspA has seven repeats which are similar to mucus-binding domains. HT29 cells produce mucins upon differentiation, mainly the exported MUC2 protein and small amounts of the epithelium-associated MUC3 protein, when cultured in a glucose-containing medium (36). Thus, it is plausible that the interaction of LspA with epithelial cells is through its binding to cell-associated mucins such as MUC3. The human intestine is covered in a layer of mucus, and the ability to bind specifically to mucus would be a desirable characteristic for colonization. It remains to be elucidated whether Caco C2 cells produce MUC3 on their surfaces, which could explain the reduced binding of UCC118/pORI19::*lspA* to Caco C2 cells. However, the recent inactivation of a gene encoding a mucus-binding protein in *L. acidophilus* revealed a significant reduction in adhesion to Caco-2 cells (12), but it was proposed that the specific interaction of this protein with this cell line is other than by binding to mucus, since no mucus production was detected.

The combination of genome-wide bioinformatic analysis and functional characterization has proved productive in this and other recent studies of *Lactobacillus* host interaction proteins (12, 47). Buck and colleagues recently analyzed the adhesion to Caco-2 cells of five targeted gene knockout strains of *L. acidophilus* NCFM (12), and they reported higher levels of adhesion reduction for mutants lacking FbpA, Mub, and SlpA. LspA, LspB, LspC, and LspD exhibit only 13.5%, 12.5%,

12.8%, and 13.6% overall identity with LBA1633, SlpA, SlpA, and FbnA of *L. acidophilus* NCFM, respectively (best reciprocal BLAST hits), and these values drop to background when the signal peptides and anchor domains are removed. Buck et al. (12) reported that inactivation of a single mucus-binding protein (LBA1392) of *L. acidophilus* NCFM reduced the adhesion to Caco-2 cells by 65%, whereas the largest reduction in adhesion that we recorded was 50%, for the adhesion of UCC118Δ*srt*A to HT29 cells. Differences in the methodologies of assays and in cell culture conditions likely combine to make comparisons of adhesion level reductions difficult in absolute terms. We noted that the PCR-based assay generally detected larger numbers of adherent bacteria than the viable count assay, probably because clumping of bacteria or aggregation with residual cell membrane in the latter assay reduced the countable CFU. We also noted generally lower reductions of adhesion levels for mutants tested against Caco-2 cells, which might indicate more abundant receptors for sortase-independent adhesins in this cell line. Clearly, however, different sortase-dependent proteins are important adhesins in *L. acidophilus* and *L. salivarius* UCC118, and our findings also suggest that sortase-independent cell surface proteins in *L. salivarius* UCC118 have a significant contribution. Experiments are in progress to identify these proteins. An improved understanding of epithelial cell adhesion mechanisms in probiotic bacteria will allow for future strain improvement or informed strain selection.

ACKNOWLEDGMENTS

REFERENCES

1. Ahrne, S., S. Nobaek, B. Jeppsson, I. Adlerberth, A. E. Wold, and G. Molin. 1998. The normal *Lactobacillus* flora of healthy human rectal and oral mucosa. J. Appl. Microbiol. **85:**88–94.
2. Altermann, E., W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer. 2005. Complete genome sequence of the probiotic lactic acid bacterium Lactobacillus acidophilus NCFM. Proc. Natl. Acad. Sci. USA **102:**3906–3912.
3. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. Barnett, T. C., A. R. Patel, and J. R. Scott. 2004. A novel sortase, SrtC2, from *Streptococcus pyogenes* anchors a surface protein containing a QVPTGV motif to the cell wall. J. Bacteriol. **186:**5865–5875.
5. Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy. 2004. The Pfam protein families database. Nucleic Acids Res. **32:**D138–D141.
6. Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
7. Bierne, H., S. K. Mazmanian, M. Trost, M. G. Pucciarelli, G. Liu, P. Dehoux, L. Jansch, F. Garcia-del Portillo, O. Schneewind, and P. Cossart. 2002. Inactivation of the srtA gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. Mol. Microbiol. **43:**869–881.
8. Boekhorst, J., M. W. de Been, M. Kleerebezem, and R. J. Siezen. 2005. Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. J. Bacteriol. **187:**4928–4934.
9. Bolken, T. C., C. A. Franke, K. F. Jones, G. O. Zeller, C. H. Jones, E. K. Dutton, and D. E. Hruby. 2001. Inactivation of the srtA gene in *Streptococcus gordonii* inhibits cell wall anchoring of surface proteins and decreases in vitro and in vivo adhesion. Infect. Immun. **69:**75–80.
10. Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek,

N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. **22:**1554–1558.
11. Bron, P. A., C. Grangette, A. Mercenier, W. M. de Vos, and M. Kleerebezem. 2004. Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. J. Bacteriol. **186:**5721–5729.
12. Buck, B. L., E. Altermann, T. Svingerud, and T. R. Klaenhammer. 2005. Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. Appl. Environ. Microbiol. **71:**8344–8351.
13. Chaillou, S., M. C. Champomier-Verges, M. Cornet, A. M. Crutz-Le Coq, A. M. Dudez, V. Martin, S. Beaufils, E. Darbon-Rongere, R. Bossy, V. Loux, and M. Zagorec. 2005. The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. Nat. Biotechnol. **23:**1527–1533.
14. Claesson, M. J., Y. Li, S. Leahy, C. Canchaya, J. P. van Pijkeren, A. M. Cerdeño-Tárraga, J. Parkhill, S. Flynn, G. C. O'Sullivan, J. K. Collins, D. Higgins, F. Shanahan, G. Fitzgerald, D. van Sinderen, and P. W. O'Toole. 2006. Multi-replicon genome architecture of *Lactobacillus salivarius*. Proc. Natl. Acad. Sci. USA, **103:**6718–6723.
15. De Boever, E. H., and D. B. Clewell. 2001. The *Enterococcus faecalis* pheromone-responsive plasmid pAM373 does not encode an entry exclusion function. Plasmid **45:**57–60.
16. de Man, J. C., M. Rogosa, and M. E. Sharpe. 1960. A medium for the cultivation of lactobacilli. J. Appl. Bacteriol. **23:**130–135.
17. Demuth, D. R., C. A. Davis, A. M. Corner, R. J. Lamont, P. S. Leboy, and D. Malamud. 1988. Cloning and expression of a *Streptococcus sanguis* surface antigen that interacts with a human salivary agglutinin. Infect. Immun. **56:** 2484–2490.
18. Dunne, C., P. Kelly, S. O'Halloran, D. Soden, M. Bennett, A. von Wright, T. Vilpponen-Salmela, B. Kiely, L. O'Mahony, J. K. Collins, G. C. O'Sullivan, and F. Shanahan. 2004. Mechanisms of adherence of a probiotic *Lactobacillus* strain during and after in vivo assessment in ulcerative colitis patients. Microb. Ecol. Health Dis. **16:**96–104.
19. Dunne, C., L. Murphy, S. Flynn, L. O'Mahony, S. O'Halloran, M. Feeney, D. Morrissey, G. Thornton, G. Fitzgerald, C. Daly, B. Kiely, E. M. Quigley, G. C. O'Sullivan, F. Shanahan, and J. K. Collins. 1999. Probiotics: from myth to reality. Demonstration of functionality in animal models of disease and in human clinical trials. Antonie Leeuwenhoek **76:**279–292.
20. Dunne, C., L. O'Mahony, L. Murphy, G. Thornton, D. Morrissey, S. O'Halloran, M. Feeney, S. Flynn, G. Fitzgerald, C. Daly, B. Kiely, G. C. O'Sullivan, F. Shanahan, and J. K. Collins. 2001. *In vitro* selection criteria for probiotic bacteria of human origin: correlation with *in vivo* findings. Am. J. Clin. Nutr. **73:**386S–392S.
21. Fischetti, V. A., V. Pancholi, and O. Schneewind. 1990. Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram positive cocci. Mol. Microbiol. **4:**1603–1605.
22. Flynn, S. 2001. Molecular characterisation of bacteriocin producing genes and plasmid encoded functions of the probiotic strain *Lactobacillus salivarius* subsp. *salivarius* UCC118. Ph.D. thesis. University College Cork, Cork, Ireland.
23. Flynn, S., D. van Sinderen, G. M. Thornton, H. Holo, I. F. Nes, and J. K. Collins. 2002. Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118. Microbiology **148:** 973–984.
24. Fuller, R. 1989. Probiotics in man and animals. J. Appl. Bacteriol. **66:**365–378.
25. Garcia, E., J. L. Garcia, P. Garcia, A. Arraras, J. M. Sanchez-Puelles, and R. Lopez. 1988. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. Proc. Natl. Acad. Sci. USA **85:**914–918.
26. Gasson, M. J. 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. J. Bacteriol. **154:**1–9.
27. Heger, A., and L. Holm. 2000. Rapid automatic detection and alignment of repeats in protein sequences. Proteins **41:**224–237.
28. Hudault, S., V. Lievin, M. F. Bernet-Camard, and A. L. Servin. 1997. Antagonistic activity exerted in vitro and in vivo by *Lactobacillus casei* (strain GG) against *Salmonella typhimurium* C5 infection. Appl. Environ. Microbiol. **63:**513–518.
29. Jonsson, I. M., S. K. Mazmanian, O. Schneewind, T. Bremell, and A. Tarkowski. 2003. The role of *Staphylococcus aureus* sortase A and sortase B in murine arthritis. Microbes Infect. **5:**775–780.
30. Kharat, A. S., and A. Tomasz. 2003. Inactivation of the srtA gene affects localization of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells in vitro. Infect. Immun. **71:**2758–2765.
31. Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA **100:**1990–1995.

32. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.

33. **Lalioui, L., E. Pellegrini, S. Dramsi, M. Baptista, N. Bourgeois, F. Doucet-Populaire, C. Rusniok, M. Zouine, P. Glaser, F. Kunst, C. Poyart, and P. Trieu-Cuot.** 2005. The SrtA sortase of *Streptococcus agalactiae* is required for cell wall anchoring of proteins containing the LPXTG motif, for adhesion to epithelial cells, and for colonization of the mouse intestine. Infect. Immun. **73:**3342–3350.

34. **Law, J., G. Buist, A. Haandrikman, J. Kok, G. Venema, and K. Leenhouts.** 1995. A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. J. Bacteriol. **177:**7011–7018.

35. **Leenhouts, K., A. Bolhuis, G. Venema, and J. Kok.** 1998. Construction of a food-grade multiple-copy integration system for *Lactococcus lactis*. Appl. Microbiol. Biotechnol. **49:**417–423.

36. **Mack, D. R., and M. A. Hollingsworth.** 1994. Alteration in expression of MUC2 and MUC3 mRNA levels in HT29 colonic carcinoma cells. Biochem. Biophys. Res. Commun. **199:**1012–1018.

37. **Maguin, E., P. Duwat, T. Hege, D. Ehrlich, and A. Gruss.** 1992. New thermosensitive plasmid for gram-positive bacteria. J. Bacteriol. **174:**5633–5638.

38. **Marino, M., M. Banerjee, R. Jonquieres, P. Cossart, and P. Ghosh.** 2002. GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands. EMBO J. **21:**5623–5634.

39. **Mattila-Sandholm, T., S. Blum, J. K. Collins, R. Crittenden, W. M. de Vos, C. Dunne, R. Fondén, G. Grenov, E. Isolauri, B. Kiely, P. Marteau, L. Morelli, A. Ouwehand, R. Reniero, M. Saarela, S. Salminen, M. Saxelin, E. Schiffrin, F. Shanahan, E. Vaughan, and A. von Wright.** 1999. Probiotics: towards demonstrating efficiency. Trends Food Sci. Technol. **10:**393–399.

40. **Mazmanian, S. K., H. Ton-That, K. Su, and O. Schneewind.** 2002. An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. Proc. Natl. Acad. Sci. USA **99:**2293–2298.

41. **Navarre, W. W., and O. Schneewind.** 1994. Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. Mol. Microbiol. **14:**115–121.

42. **Navarre, W. W., and O. Schneewind.** 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol. Mol. Biol. Rev. **63:**174–229.

43. **Pastar, I., I. Tonic, N. Golic, M. Kojic, R. van Kranenburg, M. Kleerebezem, L. Topisirovic, and G. Jovanovic.** 2003. Identification and genetic characterization of a novel proteinase, PrtR, from the human isolate *Lactobacillus rhamnosus* BGT10. Appl. Environ. Microbiol. **69:**5802–5811.

44. **Patti, J. M., H. Jonsson, B. Guss, L. M. Switalski, K. Wiberg, M. Lindberg, and M. Hook.** 1992. Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesin. J. Biol. Chem. **267:**4766–4772.

45. **Paulsen, I. T., L. Banerjei, G. S. Myers, K. E. Nelson, R. Seshadri, T. D. Read, D. E. Fouts, J. A. Eisen, S. R. Gill, J. F. Heidelberg, H. Tettelin, R. J. Dodson, L. Umayam, L. Brinkac, M. Beanan, S. Daugherty, R. T. DeBoy, S. Durkin, J. Kolonay, R. Madupu, W. Nelson, J. Vamathevan, B. Tran, J. Upton, T. Hansen, J. Shetty, H. Khouri, T. Utterback, D. Radune, K. A. Ketchum, B. A. Dougherty, and C. M. Fraser.** 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science **299:**2071–2074.

46. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29:**2002–2007.

47. **Pretzer, G., J. Snel, D. Molenaar, A. Wiersma, P. A. Bron, J. Lambert, W. M. de Vos, R. van der Meer, M. A. Smits, and M. Kleerebezem.** 2005. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. J. Bacteriol. **187:**6128–6136.

48. **Pridmore, R. D., B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni, and M. A. Schell.** 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc. Natl. Acad. Sci. USA **101:**2512–2517.

49. **Roos, S., and H. Jonsson.** 2002. A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. Microbiology **148:**433–442.

50. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. **132:**365–386.

51. **Serror, P., T. Sasaki, S. D. Ehrlich, and E. Maguin.** 2002. Electrotransformation of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis* with various plasmids. Appl. Environ. Microbiol. **68:**46–52.

52. **Sheil, B., J. McCarthy, L. O'Mahony, M. W. Bennett, P. Ryan, J. J. Fitzgibbon, B. Kiely, J. K. Collins, and F. Shanahan.** 2004. Is the mucosal route of administration essential for probiotic function? Subcutaneous administration is associated with attenuation of murine colitis and arthritis. Gut **53:**694–700.

53. **Sonnhammer, E. L., and R. Durbin.** 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:**GC1–GC10.

54. **Stalhammar-Carlemalm, M., T. Areschoug, C. Larsson, and G. Lindahl.** 1999. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. Mol. Microbiol. **33:**208–219.

55. **Steen, A., G. Buist, K. J. Leenhouts, M. El Khattabi, F. Grijpstra, A. L. Zomer, G. Venema, O. P. Kuipers, and J. Kok.** 2003. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. J. Biol. Chem. **278:**23874–23881.

56. **Sturme, M. H., J. Nakayama, D. Molenaar, Y. Murakami, R. Kunugi, T. Fujii, E. E. Vaughan, M. Kleerebezem, and W. M. de Vos.** 2005. An *agr*-like two-component regulatory system in *Lactobacillus plantarum* is involved in production of a novel cyclic peptide and regulation of adherence. J. Bacteriol. **187:**5224–5235.

57. **Sutcliffe, I. C., and D. J. Harrington.** 2002. Pattern searches for the identification of putative lipoprotein genes in gram-positive bacterial genomes. Microbiology **148:**2065–2077.

58. **Terzaghi, B. E., and W. E. Sandine.** 1975. Improved medium for lactic streptococci and their bacteriophages. Appl. Microbiol. **29:**807–813.

59. **Ton-That, H., G. Liu, S. K. Mazmanian, K. F. Faull, and O. Schneewind.** 1999. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. Proc. Natl. Acad. Sci. USA **96:**12424–12429.

60. **Ventura, M., C. Canchaya, V. Bernini, E. Altermann, R. Barrangou, S. McGrath, M. J. Claesson, Y. Li, S. Leahy, C. D. Walker, R. Zink, E. Neviani, J. Steele, J. Broadbent, T. Klaenhammer, G. Fitzgerald, P. W. O'Toole, and D. van Sinderen.** 2006. Comparative genomics and transcriptional analysis of prophages identified in the genomes of *Lactobacillus gasseri*, *Lactobacillus salivarius*, and *Lactobacillus casei*. Appl. Environ. Microbiol., **72:**3130–3146.